

Index

- accuracy, 11
- active learning, 16
- ADF, *see* test of stationarity
- AFE, *see* automated feature engineering
- affine transformations, 222, 223
- area under the ROC curve, *see* AUC-ROC
- artificial neural network, *see* neural network
- attribute, *see* raw data
- AUC-ROC, 12
- Augmented Dickey Fuller, *see* test of stationarity
- autocorrelation, 121
- autoencoder, 134
 - variational, 135
- automated feature engineering, 130
- AutoML, 32, 131

- backpropagation through time, *see* recurrent neural network
- backpropagation-through-time, 130
- bag of words, 52, 196
- Bayesian information criterion, *see* BIC
- bias, 12
- BIC, *see* feature utility
- bin counting, *see* target rate encoding
- binarize, *see* thresholding
- binning, 44
 - equal frequency intervals, 45
 - equal interval width, 45
- blob tracking, 237
- bootstrap, 19
- bottleneck, 134
- BoW, *see* bag of words
- Box-Cox transformation, *see* squashing function

- BPTT, *see* backpropagation through time
- bucketing, *see* binning

- camera calibration, 42
- Cartesian product, 63
- categorical encoding, *see* target rate encoding
- censored data, 55
- censored values, *see* imputation
- centering, *see* feature centering
- CFAR, *see* target rate encoding
- changing coordinate system, 64
- chartless calibration, 230
- clustering, 46, 104
 - canopy, 104
 - k-means, 46, 169
- coalesced category, 66
- collisions, 100
- common sense knowledge, 66
- complex features, *see also* feature, 30
 - dates, 71
 - decomposition, 70
 - locations, 72
 - pivot, 72
 - string splitting, 71
 - times, 71
- computable features, *see* feature
- confirmation set, *see* unseen data
- confusion table, 82
- contingency table, 10
- convolution, 131, 219
- convolution kernel, 131, 175, 219
- corner detection, 226
- corpora, 106, 210
- correct firing rate, *see* target rate encoding

- correct first attempt rate, *see* target rate
- encoding
- correlation coefficient, 121
- correlation-based features, 56
- correlogram, 177
- covariance, 121
- covariance matrix, 38
- cross-validation, 13
- curse of dimensionality, 15
- data cleaning, 6
- data pipeline, 6
- data release schedule, 19
- data velocity, 166
- DBpedia, 142
- de-pivot, *see* pivot
- decorrelation, *see also* feature normalization, 38
- deep learning, 127
- delta features, 56
- descriptive features, 52
- descriptive statistics, 50
- differential features, 171
- dimensionality reduction, 99, 134
 - clustering, 104
 - embeddings, 106
 - feature hashing, 100, 205
 - latent Dirichlet allocation, 103
 - LDA, *see* latent Dirichlet allocation, *see also* linear discriminant analysis
 - linear discriminant analysis, 106
 - PCA, *see* SVD
 - pLSA, *see* latent Dirichlet allocation
 - probabilistic latent semantic analysis, *see* pLSA
 - random projection, 101
 - SVD, 102
 - t-SNE, 107, 206
- discrete binning, *see* binning
- discretization, 43, 63, 155
 - adaptive quantizer, 48
 - ChiMerge, 48
- discretization error, 43
- discretization, supervised, 43
- discretization, unsupervised, 43
- distributed representations, *see* embeddings
- distributional hypothesis, 106
- DL, *see* deep learning
- domain expertise, 2, 8, 26, 138, 186
- domain knowledge, 35, 66, 125
- domain modelling, 25
- dropout, 128
- dummy variables, *see* one hot encoding
- EDA, 21, 144, 167, 177, 190, 215
- edge detection, 230
- elastic net, *see* regularization
- EM algorithm, 47
- EMA, *see* exponential moving average
- embedded feature selection, 94
- embeddings, *see also* dimensionality reduction, 106
 - global, 108
 - local, 109, 206
 - Word2Vec, *see* local
- engineered feature, *see* computable feature
- error analysis, 22, 152, 153, 172, 199, 202, 218, 221, 223, 224, 226
- error models, 120
- expectation-maximization, *see* EM algorithm
- exploratory data analysis, *see* EDA
- exponential moving average, 173
- external data, 66, 73
- f-measure, 11
- fake variables, *see* random feature
- false colour, 216
- false negatives, 11
- false positives, 11
- feature, 6
 - binary, 28
 - categorical, 28
 - complex, 30, 64
 - computable, 60
 - construction, 27
 - continuous, 28
 - discrete, 28
 - drill-down, 61
 - evolution, 15
 - importance, 89
 - imputation, *see* imputation
 - indicator, 64, 113
 - learning, 131
 - normalization, *see* feature normalization
 - perturbation, 134, 159
 - template, 30
- feature centering, *see also* feature normalization, 37
- feature drill-down, 61
- feature engineering
 - definition, 7
 - process, 19, 143
 - unsupervised, 134

feature explosion, 25
 feature ideation, 25
 feature map, *see also* kernel methods, 74, 77
 feature normalization, 35
 feature scaling, *see also* feature normalization, 36
 feature selection, 80
 backward, 90
 blacklisting features, 94
 embedded, 94
 filter, 90
 forward, 90
 greedy, 90
 greedy methods, 90
 implicit, 95
 stability, 93
 wrapper methods, 88, 154
 feature set
 ablation study, 91, 154
 incremental construction, 90
 stability, 159
 feature utility, 82
 AIC, 93
 ANOVA, 86, 88
 chi square, 83
 firing rate, 83
 FPR, 83
 Gini Impurity, 85
 information gain, 85
 maximum likelihood estimator, 88
 MLE, *see* maximum likelihood estimator
 mutual information, 84, 156
 Pearson correlation coefficient, 83
 TPR, 83
 feature/target transformations, 62, 154
 Featuretools, 32, 131
 featurization, 6
 featurizer, 26
 fisher score, 87
 Fisher's criterion, 88
 frequency domain, 41
 Frobenious norm, 105

 Gaussian blur, 219
 Gaussian kernel, *see* kernel
 genetic programming, 133
 GeoNames, 143
 gradient descent, 127
 Graphviz, 144

hashing function, 100
 held out data, 18
 histogram, 50, 114, 224
 histogram backprojection, 237
 histogram of oriented gradients,
 see HOG
 HOG, 227

 ICA, 105
 implicit feature selection, 94
 implicit feature space, *see* kernel methods
 imputation, 67, 152
 categorical, 69
 representative value, 69
 timestamped, 69, 170
 indicator feature, *see* feature, binary
 instance-based engineering, 124, 222
 instances, 6
 interaction variable, *see* computable feature

 k-NN algorithm, 70
 kernel
 Gaussian, *see* RBF
 polynomial, 76
 RBF, 77
 kernel methods, 73
 kernel trick, *see* kernel methods
 key features, 173
 key frames, 236
 Kullback–Leibler divergence, 108
 kurtosis, 53

 lab notebook, 18
 lag features, 171
 lags, 121
 LASSO, *see* regularization
 LDA, *see* dimensionality reduction
 learning
 active, *see* active learning
 reinforcement, *see* reinforcement learning
 semi-supervised, *see* semi-supervised learning
 supervised, *see* supervised learning
 unsupervised, *see* unsupervised learning
 leave one out encoding, *see* target rate encoding
 lemma, 118
 lemmatizer, 201
 local feature detectors, 225
 log book, *see* lab notebook
 low rank approximation, 105

- Mahalanobis distance, 39
- Mahalanobis transformation, *see* ZCA
- Manhattan distance, *see* norm
- Markov assumption, 170
- Markov features, 171, 180
- Markov processes, 170
- MDL, *see* minimum description length
- mean normalized values, *see* feature centering
- mean squared error, *see* MSE
- melting, *see* pivot
- Mercer's theorem, 75
- mini-batches, 127
- minimum description length, 95
- minimum spanning tree, *see* spanning tree
- missing data indicator feature, 68, 159, 171
- missing values, 55, *see* imputation, 149
- most general tree, 116
- MSE, 12
- MST, *see* spanning tree
- multinomial theorem, 77
- mutual information, 84

- n-grams, 115
- neural network, 127
 - recurrent, 129
- NMF, *see* non-negative matrix factorization
- non-negative matrix factorization, 105, 245
- norm
 - Euclidean, 37
 - L1, 37, 96
 - L2, 37, 96
 - Manhattan distance, *see* L1
- normal form, *see* tidy data
- nuisance variations, 36, 51, 114, 193, 222

- one-hot-encoding, 64, 128, 152
- OOF, *see* out-of-fold
- operator palette, 61
- opportunistic imputation, 242
- out-of-fold, 14, 70, 158, 162
- outlier
 - score, 55
- outliers, 54, 150
 - detection, 55
 - king effect, 55
- overfitting, 14

- padding, 114
- parent-node pairs, 117
- part-of-speech, 118

- PCA
 - see* dimensionality reduction, 102
 - PCA decomposition, 38
 - perfect hash, 100
 - periodogram, 185
 - pivot, 21, 72
 - polynomial features, 63
 - polynomial kernel, *see* kernel
 - polysemy, 101
 - positive predictive value, *see* precision
 - PPV, *see* positive predictive value
 - precision, 11
 - principal component analysis
 - see* PCA, 102
 - probe features, *see* random feature

- random feature, 91
- raw data, 6
- RBF kernel, *see* kernel
- RBM, *see* restricted Boltzmann machine
- RDF, 142
- recall, 11
- rectified linear units, *see* ReLU
- recurrent neural network, *see also* neural networks, 129
- recursive feature elimination, 91
- regularization
 - elastic net, 97
 - L1, 96
 - L2, 96
 - LASSO, *see* L1
 - ridge regression, *see* L2
- regularization techniques, 95
- reinforcement learning, 16
- ReLU, 127
- representation
 - graphs, 118, 143
 - lists, 114, 147
 - sets, 113, 147
 - trees, 116
- reproducible builds, 18
- residuals, 124
 - training on, 183
- restricted Boltzmann machine, 134
- RFE, *see* recursive feature elimination
- ridge regression, *see* regularization
- RMSE, 12, 150
- RNN, 124, *see* recurrent neural network

- scaling to unit length, *see* feature normalization

Schwarz criterion, *see* BIC
 semantic graph, 142
 semi-supervised learning, 16
 sensitivity, *see* recall
 shadow features, *see* random feature
 shift-reduce encoding, 118
 sigmoid function, *see* squashing function
 simple moving average, 173
 single feature transformations, 61
 singular value decomposition, *see*
 dimensionality reduction
 skewness, 53
 skip n-grams, 115, 204
 Slutsky-Yule effect, 173
 SMA, *see* simple moving average
 smoothing, *see also* feature normalization, 40
 ELE, 41
 Lagrangian, 41
 simple, 41
 simple Good-Turing, 41
 softmax layer, 127
 spanning tree, 119
 sparse coding, 105
 spectral norm, 105
 sphering transformation, *see* whitening
 squashing function, 62
 standardization, 36, *see also* feature
 normalization, 38
 stationarity, 122
 stemmer, 201
 stop words, 94, 200
 stratified cross-validation, 175
 string splitting, 71
 SUBSET, 89
 supervised learning, 16
 support vector machines, *see* SVM
 SVD, *see* dimensionality reduction
 SVM, *see* kernel methods
 synthetic instance, 46

t-SNE, *see* dimensionality reduction
 target class, 6
 target leak, 14, 145, 193
 target rate encoding, 65
 target value, *see* target class

temporal smoothing, 173
 test of stationarity, *see also* stationarity, 123
 TF-IDF, 42
 thought features, 25
 thresholding, *see also* discretization, 45, *see*
 winsorising
 tidy data, 73
 tied weights, 129
 time series, 176
 cycle, 122
 seasonal, 122
 trend, 121
 TPR, *see* true positive rate
 true positive rate, *see* recall
 type I errors, *see* false positives
 type II errors, *see* false negatives

unknown values, *see* imputation
 unobserved events, 40
 unseen data, 15
 unstacking, *see* pivot
 unsupervised learning, 16, 46, 99
 utility metrics, 12

VAE, *see* autoencoder
 vanishing gradients, 129
 variable, *see* feature
 variable-length feature vectors, 112
 variance, 12
 variance scaling, *see* feature normalization
 vector space model, *see* embeddings

whitening, 39, 220
 PCA, 39
 WikiCities, 142
 winsorising, *see* thresholding, 63
 word embedding, *see* embeddings
 Word2Vec, *see* embeddings
 WordNet, 104
 wrapper method, 88

Xavier initialization, 127

z-scores, 57
 ZCA, *see* whitening